Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 46 (2006) 1488–1492

Vision Research

www.elsevier.com/locate/visres

Monocular texture segmentation and proto-rivalry

Joshua A. Solomon *, Adrian John, Michael J. Morgan

Department of Optometry and Visual Science, City University, London EC1V 0HB, UK

Received 11 April 2005; received in revised form 28 June 2005

Abstract

When the right eye's target is the left eye's distracter and vice versa, orientation-defined search is impossible unless, as we show here, the elements are close together. More than 1 s was required to find inverse-cyclopean texture boundaries when elements were arranged on a 16×16 grid. Less than 250 ms was required for a 24×24 grid covering the same area. The conventional view is that binocular rivalry requires at least 200 ms to develop, but our results suggest a more rapid access to monocular signals. We call this rapid form of access "proto-rivalry."

Keywords: Binocular; Dichoptic; Cyclopean; Search

1. Introduction

Binocular rivalry refers to the alternation in consciousness of two images, presented separately to the two eyes. Ever since Hering (1874) noted that brief dichoptic displays of orthogonal lines appeared as a uniform grid, conventional wisdom (Dawson, 1913; Howard, 2002) has dictated that rivalry requires at least 200 ms to develop.

However, it has recently become clear that the pre-rivalrous experience is not simply binocular fusion. For instance, we now know that observers can easily discriminate between plaids constructed from dichoptically presented component gratings and those in which the two components are optically fused (Blake, Yang, & Wilson, 1991; Georgeson & Meese, 1997). One recent report even suggests that (although their apparent direction is consistent with fusion) drifting, dichoptic plaids do not appear fused at all.[1]

Our interest in pre-rivalrous experiences was stimulated by Kolb and Braun (1995), who demonstrated that orientation-defined texture boundaries, which should be erased by binocular fusion, nonetheless can be located in 100-ms displays. Similar "inverse-cyclopean" (Julesz, 1971) texture boundaries are shown in Fig. 1A. When presented dichoptically, the two "target" plaids appear no different from any of the other plaids, yet some cue allows observers to segregate them from their neighbours.

We have confirmed inverse-cyclopean texture segmentation, and suggested a class of monocular neurones sensitive to orientation contrast (Morgan, Mason, & Solomon, 1997; Solomon & Morgan, 1999). On the other hand, Wolfe and Franzel (1988) found that orientation-defined search was impossible when the right eye's target was the left eye's distracter, and vice versa. Although there were several differences between Kolb and Braun's (1995) and Wolfe and Franzel's displays, we suspected that the critical one was element density. We also wondered whether rapid texture segmentation could occur when the monocular signals did not contain elements competing for the same position in the visual field.

* Corresponding author. Tel.: +44 20 7040 0192; fax: +44 20 7040 0182.

E-mail address: j.a.solomon@city.ac.uk (J.A. Solomon).

[1] Andrews and Blakemore (1999) report seeing just one component.

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.07.002

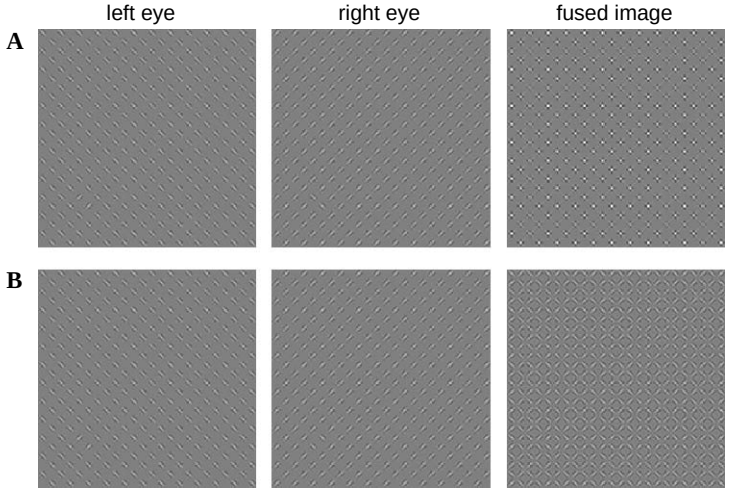

Fig. 1. (A) "Competitive" and (B) "non-competitive" displays. Dense arrays (i.e., 24×24 grids) are illustrated. Sparser arrays (i.e., 20×20 and 16×16 grids) contained fewer (but identical) Gabor patterns, distributed over the same area.

2. Experiment 1

2.1. Methods

The three authors served as observers. They all had previous experience with inverse-cyclopean texture segmentation. JAS and MJM wore corrective spectacles.

The 37.5-cd/m^2 display was viewed in a dark room from 33 cm. Display resolution was 13 pixels/deg. Ferro-optical goggles (Cambridge Research Systems) were used to separate the left- and right-eye's images. The frame rate of the monitor was 140 Hz (i.e., 70 Hz for each eye).

All texture elements were maximum contrast, cosine-phase Gabor patterns (i.e., they all had a central white stripe) whose wavelength, spread and orientation were $\lambda = 0.28°$, $\sigma = 0.14°$ and $\pm 45°$ w.r.t. vertical, respectively. The elements in each display were arranged on a notional checkerboard. In the competitive displays (e.g., Fig. 1A), both eyes' elements occupied only the red or black squares. In the non-competitive displays (e.g., Fig. 1B), one eye's elements occupied the red squares and the other eye's elements occupied the black squares. The notional checkerboard contained either 24×24 squares (192×192 pixels), 20×20 squares (200×200 pixels) or 16×16 squares (192×192 pixels). Dichoptically presented nonius lines were constantly present at fixation. Each display contained a target (four Gabor patterns) in one quadrant of the display.

The orientation (randomly selected on each trial) of right-eye elements within the target region matched that of left-eye elements outside the target region and vice versa. Target positions were chosen to equate retinal eccentricity across densities. For example, in the 24×24 grid, the target region contained squares in the fifth and sixth rows and the fifth and sixth columns; in the 20×20 grid, it contained squares in the fourth and fifth rows and columns and in the 16×16 grid it contained squares in the third and fourth rows and columns.

Each experimental session contained 400 trials with a single display density. Half of these trials were dichoptic. In the other half, all Gabors were displayed to both eyes at one-half maximum contrast. In both of these conditions, trials were equally divided between the competitive and non-competitive conditions and five exposure durations, logarithmically spaced between 0.014 and 1.2 s. The number of sessions completed by each observer is shown in Table 1.

The observer's task was to indicate whether the target was on the left or right side of the display. Frequencies correct ψ_{2AFC}, were maximum-likelihood fit with a log-normal distribution

Table 1
Procedural and statistical details, Experiment 1

	Density	N	P value, competitive	P value, non-competitive
MJM	24×24	4	0.0143	0.0116
	20×20	2	0.0109	0.0032
	16×16	2		
JAS	24×24	3	0.0027	0.0452
	20×20	1	0.2881	0.7354
	16×16	1		
AJ	24×24	6	0.0003	1.0×10^{-5}
	20×20	3	0.1868	0.5564
	16×16	2		

N = number of sessions; P value $= 1 - \chi_2^2(2 \ln \Lambda)$, where Λ is the ratio between the maximum likelihood that data from two densities were generated by the same (lognormal) psychometric function (Eq. (1)) and the maximum likelihood that they were generated by different (lognormal) psychometric functions and χ_2^2 is the cumulative distribution function of a chi-square random variable with two degrees of freedom.

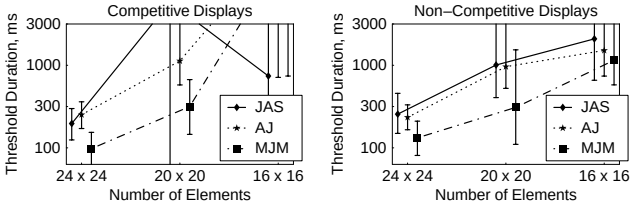

Fig. 2. Threshold exposures for segmenting dichoptically presented competitive and non-competitive displays. Error bars contain 95% confidence intervals.

$$\psi_{2\mathrm{AFC}}(t;\mu,\sigma) = 0.5 + 0.49\exp\left[-\frac{(\ln t - \ln \mu)^2}{2\sigma^2}\right] \qquad (1)$$

modified to accommodate observers' inability to respond perfectly to perfectly obvious stimuli (Wichmann & Hill, 2001). In the preceding expression, t is the exposure duration; μ and σ were allowed to vary freely. For each observer and display density, the best-fitting value for μ was considered to be threshold for texture segmentation.

2.2. Results

Threshold exposures for segmenting dichoptically displayed textures are shown in Fig. 2. Chi-square analyses (Mood, Graybill, & Boes, 1974) indicate that each observer required significantly more time to find the target when the total number of (both eye's) display elements decreased from 24×24 to 20×20 ($p < .05$, see Table 1), regardless whether the monocular signals competed for the same positions in the visual field. Neither JAS's nor AJ's threshold changed significantly when the number of elements decreased further to 16×16 ($p > .18$, see Table 1). All observers (except JAS/competitive) required a full second (i.e., 1.0 s) exposure to achieve threshold accuracy with the sparsest displays.

When the two eye's images in the non-competitive condition were combined on the video display (instead of inside the observer's visual system), more than 1 s was required to find the boundaries (all observers, all densities; data not shown). Naturally, performance did not exceed chance in the binocular (as opposed to dichoptic) competitive condition.

2.3. Discussion

We found a strong effect of density with dichoptic displays. However, under ordinary viewing conditions, orientation-defined texture segmentation in regular arrays is easy regardless of element density (Nothdurft, 1990; cf. Nothdurft, 1985). Indeed, when we removed one eye's elements, we found that all of our textures segregated perfectly (i.e., accuracy was near 100%) in less than 100 ms, the lowest threshold obtained with dichoptic displays. Experiment 2 was designed to determine whether, under ordinary viewing conditions, element density would affect orientation-defined texture segmentation, when that segmentation was not trivially easy.

Table 2
Procedural and statistical details, Experiment 2

	Density	N	P value
MJM	12×12	5	0.0001
	10×10	5	0.7390
	8×8	5	
JAS	12×12	10	0.0030
	10×10	10	0.6417
	8×8	10	
FGV	12×12	10	0.0432
	10×10	10	0.0368
	8×8	10	

Layout follows pattern of Table 1. Psychometric functions had the general form given in Eq. (2).

3. Experiment 2

3.1. Methods

Two authors served as observers. A third observer was naïve to the purpose of the experiment. All three wore corrective spectacles.

The 10.0-cd/m^2 display was viewed in a dark room from 46 cm. Display resolution was 13 pixels/deg. The frame rate of the monitor was 120 Hz.

Gabor textures were identical to those used in Experiment 1.[2] Interleaved with each video-refresh of Gabor texture was a random sample of noise, in which each pixel was drawn from a uniform distribution of luminances between 0.15 and 20.0 cd/m^2.

Each experimental session contained 40 trials with a single display density. The observer's task was to indicate which of the four quadrants contained the target. The QUEST procedure (Watson & Pelli, 1983) converged on the (threshold) exposure duration required for observers to achieve 62% accuracy.[3] The number of sessions completed by each observer is shown in Table 2.

The observer's task was to indicate which of the four quadrants contained the target. Thus Eq. (1) had to be modified to reflect a minimum probability of 0.25:

$$\psi_{4\mathrm{AFC}}(t;\mu,\sigma) = 0.25 + 0.74\exp\left[-\frac{(\ln t - \ln \mu)^2}{2\sigma^2}\right]. \qquad (2)$$

[2] Only one eye's elements were used, thus the total densities were half those of Experiment 1.

[3] QUEST assumed an underlying psychometric function having the form $\psi(t;\alpha) = 0.99 - 0.74\exp(-\sqrt{t/\alpha})$, where α is threshold (cf. Eq. (2)).

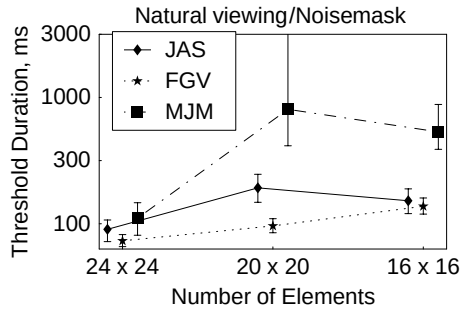

Fig. 3. Threshold exposures for segmenting noise-masked, orientation-defined textures under ordinary (binocular; not dichoptic) viewing conditions.

3.2. Results

Threshold exposures for segmenting orientation-defined textures under ordinary viewing conditions are shown in Fig. 3. Chi-square analyses (Mood et al., 1974) indicate that each observer required significantly more time to find the target when the total number of display elements decreased from 12×12 to 10×10 ($p < .05$, see Table 2). Neither JAS's nor MJM's threshold changed significantly when the number of elements decreased further to 8×8 ($p > .6$, see Table 2).

3.3. Discussion

The effect of element density on orientation-defined texture segmentation was significant (and thus conceptually consistent with orientation-energy models, e.g., Bergen & Landy, 1991), but not as strong as the effect of element density on inverse-cyclopean texture segmentation. Moreover, in order to make the two tasks comparable, we had to increase the former's difficulty with the addition of visual noise.

4. General discussion

Howard (2002) offered two explanations for rapid texture segmentation with inverse-cyclopean stimuli. One possibility was that texture borders could be revealed to binocular mechanisms simply by binocular misalignment. The results of Experiment 1 are hard evidence against this idea. First, note that our non-competitive condition was not strictly inverse cyclopean. It is possible to find the target when left- and right-eye images are optically fused (as they are in 1B, right panel). Thus, it mimics binocular misalignment in the truly inverse-cyclopean "competitive" condition. However, we found that when fusion was optical, observers required more than 1 s to locate the texture borders. All observers enjoyed a sizeable advantage when the dense textures were displayed dichoptically.

Vergence instability is a potentially more powerful cue. Howard (2002) noted that it is consistent with observers inability to segregate inverse-cyclopean textures in moving displays (p. 293). However, despite the fact that Howard was referring to our earlier work (Solomon & Morgan, 1999), we did not explicitly examine texture segregation with dynamic displays. (Rather, we examined direction discrimination, which was in fact possible in certain circumstances.) We (dichoptically) simulated vergence fluctuations with S. Tripathy. Target localization was perfect with 250-ms displays and 1.2°, non-conjugate fluctuations at 30 Hz. On the other hand, target localization was at chance levels after 1 s of (a) 0.12° fluctuations at 30 Hz and (b) 1.2° fluctuations at 2 Hz. We must concede that large vergence fluctuations could explain the basic phenomenon. Image stabilization (or, at least, vergence monitoring) with a binocular eye tracker is required to settle this issue.

Assuming no such artefacts, segmentation of inverse-cyclopean textures is consistent with a monocular texture process. It would be impossible if both eyes' stimuli were combined prior to viewing and, given the rapidity with which it is possible, it can be accomplished without binocular rivalry. Second-order filtering could occur prior to binocular combination. Monocular units are prevalent in V1, but are not absent from pre-striate areas (Baker, Grigg, & von Noorden, 1974; Burkhalter & Van Essen, 1986).

Alternatively, rapid texture segmentation could be performed by a binocular process if attenuation of monocular signals were more rapid than the 200-ms minimum implied by studies of binocular rivalry (Dawson, 1913; Howard, 2002). Observers need not be conscious of a dominant monocular signal during this period of *proto-rivalry*. Subjective fusion may be accomplished by some mechanism different from the texture processor. Another possibility is that rivalry is occurring in briefly presented patterns but at such a high rate as to be indistinguishable subjectively from fusion (a kind of low-pass flicker fusion limit for awareness). In other words, texture segregation mechanisms may have access to inputs whose small temporal separation makes them seem simultaneous. This idea is similar to the possibility that texture averaging mechanisms have access to "crowded" inputs whose small spatial separation hampers individuation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Proto-rivalry is subtly different from Kolb and Braun's (1995) original suggestion that monocular texture segmentation is a form of blindsight. Kolb and Braun concluded that observers were not conscious of inverse-cyclopean texture boundaries. Our results suggest that observers are conscious of these boundaries, but they remain unconscious of the monocular signals that contain them.

The relative ease with which (unmasked) monocular, orientation-defined textures can be segmented (compared

to inverse-cyclopean textures) is consistent with all of these explanations. For example, during proto-rivalry, monocular signals may be degraded due to their brevity and/or incomplete dominance. Furthermore, if mechanisms for texture segmentation did exist at a monocular stage of processing, they may be greatly out-numbered by similar mechanisms at binocular stages. Both would respond to monocularly presented stimuli, but only the former could respond to inverse-cyclopean stimuli.

In summary, we have found that orientation-defined texture segmentation is fastest with dense textures and this effect of density is even more striking for inverse cyclopean textures. It remains to be determined whether the critical variable is the visual angle between Gabor elements or its ratio with each element's central frequency.

References

Andrews, T. J., & Blakemore, C. (1999). Form and motion have independent access to consciousness. *Nature Neuroscience, 2*(5), 405–406.

Baker, F. H., Grigg, P., & von Noorden, G. K. (1974). Effects of visual deprivation and strabismus on the response of neurons in the visual cortex of the monkey, including studies on the striate and prestriate cortex in the normal animal. *Brain Research, 66*, 185–208.

Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing*. Cambridge, MA: MIT Press.

Blake, R., Yang, Y. D., & Wilson, H. R. (1991). On the coexistence of stereopsis and binocular rivalry. *Vision Research, 31*(7–8), 1191–1203.

Burkhalter, A., & Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *Journal of Neuroscience, 6*(8), 2327–2351.

Dawson, S. (1913). Binocular and uniocular discrimination of brightness. *British Journal of Psychology, 6*, 78–108.

Georgeson, M. A., & Meese, T. S. (1997). Perception of stationary plaids: the role of spatial filters in edge analysis. *Vision Research, 37*(23), 3255–3271.

Hering, E. (1874). Outlines of a theory of the light sense. Cambridge, MA 1964, Harvard University Press.

Howard, I. P. (2002). Seeing in depth v. 1: Basic mechanisms. (Ontario: I. Porteous).

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377*, 336–338.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics. McGraw-Hill series in probability and statistics*. New York: McGraw-Hill.

Morgan, M. J., Mason, A. J. S., & Solomon, J. S. (1997). "Blindsight" in normal observers. *Nature, 385*, 401–402.

Nothdurft, H. C. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research, 25*(12), 1957–1968.

Nothdurft, H. C. (1990). Texton segregation by associated differences in global and local luminance distribution. *Proceedings of the Royal Society of London Series B—Biological Sciences, 239*(1296), 295–320.

Parkes, L., Lund, J., Angelucci, A., Solomon, J., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*, 739–744.

Solomon, J. A., & Morgan, M. J. (1999). Dichoptically cancelled motion. *Vision Research, 39*(14), 2293–2297.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*(2), 113–120.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics, 63*(8), 1293–1313.

Wolfe, J. M., & Franzel, S. L. (1988). Binocularity and visual search. *Perception and Psychophysics, 44*(1), 81–93.